Collaboratively Assessing Information Quality on the Web

Hongwei Zhu
Information Technology & Decision Sciences, Old Dominion University, Norfolk, VA, United States., hzhu@odu.edu

Yinghua Ma
Shanghai Jiaotong University, Shanghai, China., mayinghua@gmail.com

Guiyang Su
Shanghai Jiaotong University, Shanghai, China., suguiyansuguiyang@gmail.com

Collaboratively Assessing Information Quality on the Web

Hongwei Zhu
Old Dominion University
hzhu@odu.edu

Yinghua Ma
Shanghai Jiaotong University
mayinghua@gmail.com

Guiyang Su
Shanghai Jiaotong University
suguiyansuguiyang@gmail.com

ABSTRACT

The Web has become a large repository of information with varying qualities. Many users often consume information without knowing its quality. Although automatic methods can be used to obtain measurements of certain aspects of quality, they are not reliable and cannot measure all aspects of quality. Users can detect errors and reliably assess aspects of quality that cannot be measured by automatic methods. However, there is a lack of technology support for users to record and share their feedback. This research aims to develop technologies to allow users to collaboratively assess information quality on the Web. The solution combines the capabilities of machines and humans to obtain comprehensive, reliable, and scalable measurements of information quality. In this paper, the crucial user interaction component of the solution is presented. It uses a browser plug-in to allow users to rate and annotate any Web page and share ratings and annotations with other users.

Keywords

Information quality, measurement, Web, human computer interaction.

INTRODUCTION

As the amount of information on the Web continues to grow, it is critical that we know the quality of the information in order to use it properly and effectively. However, due to the lack of scalable and reliable methods to assess *Information Quality* (IQ), *IQ metadata* (i.e., information about the quality of the information) is scarce. Existing IQ assessment methods largely fall into one of the two categories:

- Machine-based: using predefined IQ metrics to automatically assess IQ. Scalable, but not reliable.
- User-based: using surveys or other forms of manual assessment by user. Reliable, but not scalable.

Furthermore, machine-based methods ignore users' perspective about quality. This assessment is incomplete because IQ is determined by information's *fitness for use* (Wang and Strong 1996) and user feedback must be considered when assessing information quality (Orr 1998). Survey-based methods tend to produce assessment results not specific enough to facilitate quality improvement and effective use of existing information.

The two approaches complement each other, but there has been no research to combine the two approaches to exploit their respective strengths. This research will fill this gap by developing a novel *system* to allow users and the system to collaboratively assess and improve the quality of information on the Web. The human-centered, mass collaboration approach is feasible as it has been successful in solving other computing problems (Doan et al. 2010 (forthcoming)). When consuming information, users can spot errors and may indeed wish to report them (Klein 2000). In addition to the system, we will also develop quality-aware search and visualization techniques to harvest the IQ metadata collaboratively created by users and machine-based algorithms of the system.

In this paper, we discuss the overall research and present preliminary results on the development of the user interaction component that enables collection of collaborative assessments of information quality. Throughout this paper, we use "information" and "data" interchangeably.

BACKGROUND: IQ AND IQ ASSESSMENT

More than two decades of research in the emerging field of IQ has developed useful theories, methodologies, and technologies for assessing, improving, and managing the quality of various types of information (Madnick et al. 2009). The concept of IQ goes beyond accuracy. It includes more than a dozen other *dimensions* such as timeliness, completeness, consistency, interpretability, accessibility, security, to name only a few (Wang and Strong 1996). These different dimensions can be grouped into different categories. Several IQ frameworks have been developed to define and categorize various IQ dimensions (Bovee et al. 2003; Price and Shanks 2005; Stvilia et al. 2007; Wang and Strong 1996). Among various IQ management methodologies, the Total Data Quality Management methodology (Madnick and Wang 1992) is one of the most used in research and practice. It suggests that information should be treated as a product (as opposed to a by-product) and managed continuously by following the cycles of *Define*, *Measure*, *Analyze*, and *Improve* (Wang et al. 1998). Hence we use the term *Information Product* (IP) to refer to a piece of information such as a Web page or the query result of the deep Web. Existing research has attempted to identify a full spectrum of IQ issues, most users are only concerned with a very few IQ dimensions. In fact, research has shown that a user typically can only handle approximately seven concepts without being confused or overwhelmed (Miller 1956). Thus it is not effective to present too many IQ dimensions when informing users or soliciting their inputs about quality. We will take this factor into account when we design the system. We will also develop mechanisms to effectively identify the most important IQ dimensions concerned by users.

Numerous machine-based IQ assessment methods have been developed. Depending on the type of the information (e.g., structured vs. structured, centrally produced vs. socially contributed, medical domain vs. IT domain), different sets of metrics are selected and automatically assessed using different input features. Functional dependency analysis (Fan 2008) and statistical analysis (Dasu and Johnson 2003) can be used to identify various quality problems in relational and other types of structured sources. Record linkage techniques (Herzog et al. 2007) can be used to detect duplicates and inconsistencies. For textual data, various quality indicators can be used as a proxy for quality metrics. The indicators can be based on content (e.g., information-to-noise ratio), metadata (e.g., Web page's last update date), or other features (e.g., HTML syntactic correctness). Up to 26 such indicators have been used to assess the quality of online health information (Eysenbach et al. 2002) and as many as 100 features have been used to train classifiers to classify Web page quality (Mandl 2006). With the growth of social media such as Wikipedia and various discussion forums, there has been growing amount of research that focuses on assessing the quality of socially contributed contents. The algorithms are usually specific to a particular type of social media platform because they rely on certain features specific to the platform. For example, various features of user contribution and revision history have been used as quality metrics for Wikipedia articles (Adler and Alfaro 2007; Stvilia et al. 2007; Zeng et al. 2006). For discussion forums, features such as poster's membership duration and number of posts have been used as an indicator for the poster's credibility/trustworthiness (Wang et al. 2009). Quality and trustworthiness of users in social computing systems are also indicative of the quality of their contributions.

Most machine-based methods are scalable and can produce IQ metadata useful for improving the effectiveness of Web search and information retrieval. However, automatic algorithms can, at best, estimate the overall quality. They cannot reliably generate ratings along quality dimensions because the relationship between selected features and quality dimensions are usually unknown or unreliable. For example, number of edits is mapped to *authority* and article length is mapped to *completeness* for Wikipedia articles (Stvilia et al. 2007). It is debatable whether such mappings make sense. Ratings along quality dimensions are necessary for explication purposes and for the effective use of information (e.g., making trade-offs between dimensions). Furthermore, certain selected metrics may be irrelevant to users in their intended uses of the information. More importantly, machine-based methods cannot capture users' perspectives about IQ.

User-based assessment relies on user inputs collected using questionnaire surveys, ratings, or freeform comments. A systematic survey instrument (Lee et al. 2002) has been used in various organizations to assess IQ perceived by users of different roles in the information supply chain. The survey method requires significant user involvement and is often used to assess a collection of IPs as a whole, thus it is not scalable to obtain real-time IQ assessment at a fine-granularity. Minimalist approach to online voting (such as thumbs up/down and "has the article helped you") does not capture sufficient information for quality improvement purposes. Freeform feedback option is cumbersome and thus rarely used by users.

User-based methods can capture users' perspectives about IQ but are not scalable. They also lack the necessary granularity and specificity in terms of the IP (in the case of the survey method) and the IQ metadata (in the case of the simple voting method). Furthermore, the lack of user incentives often results in scarcity of useful feedback and even leads to biased and malicious feedback. We realize that these challenges require further research.

PROPOSED SOLUTION

We propose to address these deficiencies by developing a collaborative IQ system that combines the strengths of machine-based and user-based methods. A high-level view of the proposed system is presented in Figure 1. The system will have the following features and advantages:

- The system implements automatic algorithms to assess IQ dimensions that are suitable for machine processing. Instead of a predefined IQ framework with a fixed set of metrics, the framework and metrics will emerge from user-user and user-machine collaborations and continuously evolve to meet the changing needs of the user community.
- The system has interactive client components to facilitate users (and guide them when needed) to collaboratively assess IQ. Users can also annotate the IPs and provide feedback about machine produced IQ assessment.
- The system combines the assessments from users and machine to produce IQ metadata. It also harvests all forms of user feedback to improve the effectiveness of automatic assessment algorithms. Association between IQ metadata and IP is maintained by the system.
- The system uses the collaboratively created IQ metadata to provide users with value-added services such as information sharing, quality-ware search and visualization, and personalized information recommendation and filtering.
- IQ metadata helps information providers continuously improve quality. As information sources and user community evolve over time, the system evolves the IQ metadata and evaluation metrics to provide up-to-date quality assessment and to suit the varying needs of users.

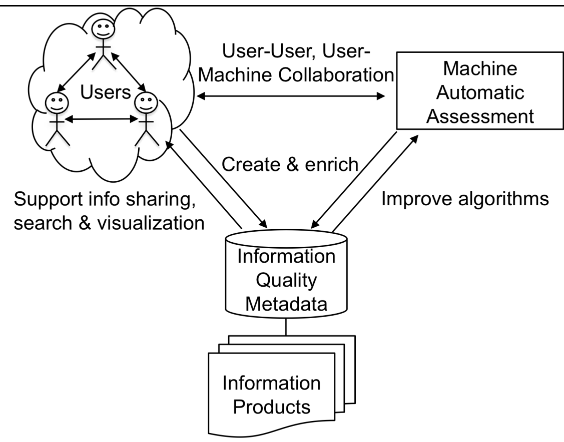

Figure 1. High-level view of the system.

The system will be designed for users to collaboratively assess the quality of any information on the Web. User involvement is crucial, but the required user effort will be minimized by distributing most tasks to the system. As illustrated in Figure 2, user task is limited to rating and annotating information using a user-friendly tool implemented as a browser plug-in. The steps for users and the system to collaboratively assess IQ are explained below (with step numbers labeled in Figure 2):

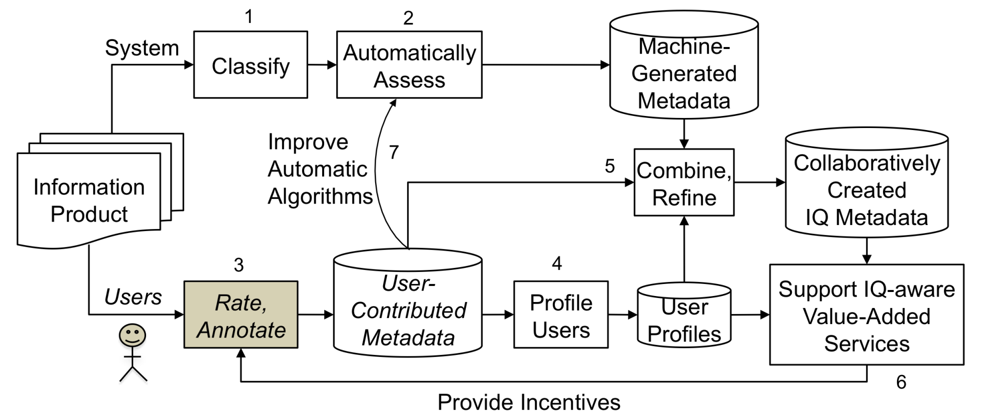

Figure 2. Process of collaboratively assessing IQ on the Web.

1. Given an information product, the system first classifies it according to *type* and *topic*.
2. Depending on the IP type, the system selects appropriate quality metrics and algorithms to assess the IP and produce "Machine-Generated Metadata".
3. Users rate and annotate the IP to produce "User-Contributed Metadata". The metadata may include quality ratings for any quality dimension (both system-suggested and user-defined), quality issues in the IP, descriptive tags, and voting of machine-generated assessment when this is presented to the user.
4. The system profiles users by analyzing their contributions to obtain "User Profiles". A user's profile indicates the user's quality/trustworthiness in providing IQ feedback. Scores are given to each user according to different topics. This is necessary because a computer scientist who gives reliable IQ assessment on computer science topics may not necessarily give reliable assessment on other topics such as medicine or finance. The fine-grained, topic-specific profiles will be built over time as the user continues to contribute more metadata.
5. The system combines machine-generated and user-contributed metadata to produce the "Collaboratively Created IQ Metadata". Most popular IQ dimensions with reliable assessment are identified. Dimensions with unreliable and potential spam assessment are also identified.
6. The system uses the IQ metadata and user profiles to provide IQ-aware and value-added services to entice users to contribute high-quality metadata. Top users are recognized to provide additional competitive incentives.
7. User-contributed metadata is also harvested to improve the automatic algorithms. For metrics and dimensions currently assessed by the algorithms, user input provides training data to enhance the performance of the algorithms. Additional metrics and dimensions suitable for automation may emerge from user feedback.
8. As more users contribute IQ metadata and data sources update their IPs, the system continuously re-assesses quality to produce update-to-date IQ metadata.

Step 3 is critical in the overall process as the effectiveness of the approach hinges on our ability to collect user inputs. This step is supported by a user interaction component of the system. We have made significant progress developing and experimenting with this component, which will be presented next.

IMPLEMENTATION OF USER INTERACTION COMPONENT

The primary design principle of the component is that it must be seamlessly integrated with user's information consuming environment. The natural choice is therefore to embed a plugin in Web browser to provide a minimally intrusive means for user to retrieve, visualize, and contribute IQ metadata. This plugin communicates with a server asynchronously to retrieve and upload IQ metadata for the information current loaded in the browser. The process of user interaction is illustrated in Figure 3.

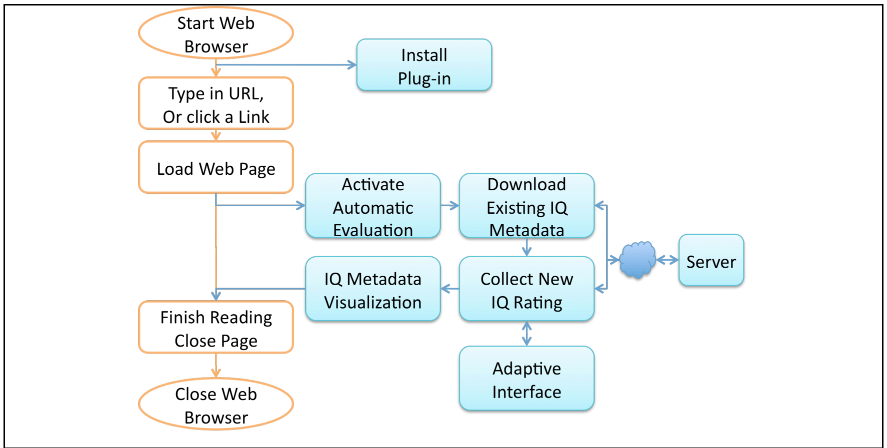

Figure 3. User interaction with the plugin and the server.

The user is only required to install the plugin and create an account once. The user will be prompted for upgrade in the future when a new version becomes available. As the browser loads a web page, the plugin asynchronously communicates with the server to download existing IQ metadata, which includes the data previously contributed by the user and by other users. In future, IQ metadata automatically generated by the system will also be retrieved. The Adaptive Interface component will determine the IQ dimensions that are most relevant to the user to dynamically update how IQ metadata is presented to the user.

We have created two prototypes of the component and are currently evaluating them. Prototype 1 (Figure 4) uses a menu bar interface. The first few menu items allow the user to supply IQ ratings. The popular/user-preferred dimensions are adaptively chosen to be visible. Other dimensions can be accessed by clicking the *More* menu item. Errors can be reported with explanations. Existing IQ metadata are retrieved from the server and displayed on the right of the toolbar. The advantage of this design is that IQ metadata preferred by the user is readily visible. Modern computer screens tend be wide. Thus the menu bar seems to take too much space vertically. This weakness can be overcome with alternative interface designs. For example, a sidebar can take advantage of wide computer displays and make more IQ metadata readily visible.

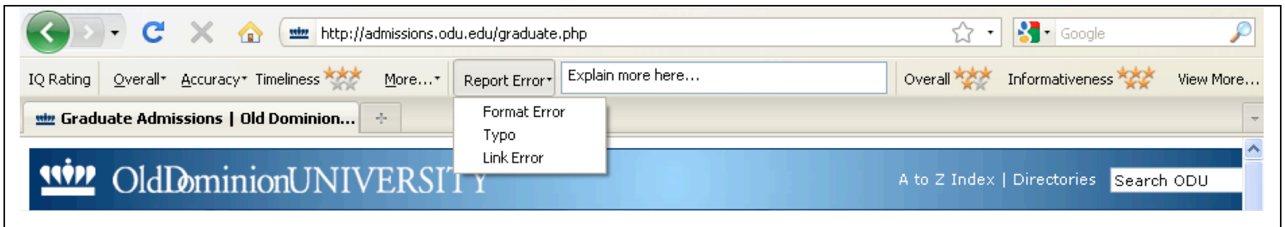

Figure 4. Prototype 1: menu bar interface.

Prototype 2 (Figure 5) overcomes the shortcoming of Prototype 1 by using an icon in browser's URL bar. Users can provide three types of feedback with the tool: (1) numeric ratings of quality, (2) comments about quality and anything else using tags and short phrases, and (3) annotation of selected texts within HTML Web pages. For the first two types of feedback, the tool is configured to provide a few dimensions, which can be customized by users. For example, for numeric ratings, the four pre-configured dimensions are Accuracy, Timeliness, Completeness, and Relevancy.

The user can view and update first two types of feedback through a pop-up window (see Figure 5) by clicking the icon. The window contains several sections. The first two sections display average scores and recently added tags about the current Web page. The next section is where the user can add or update numeric ratings of quality along selected dimensions (which are labeled as Category in the tool). Below that is where the user can add tags and short phrases. When Save button is clicked, updates will be saved on the server. To customize the categories or logout the tool, the user needs to use a configuration interface activated by clicking the Options link in the lower left corner.

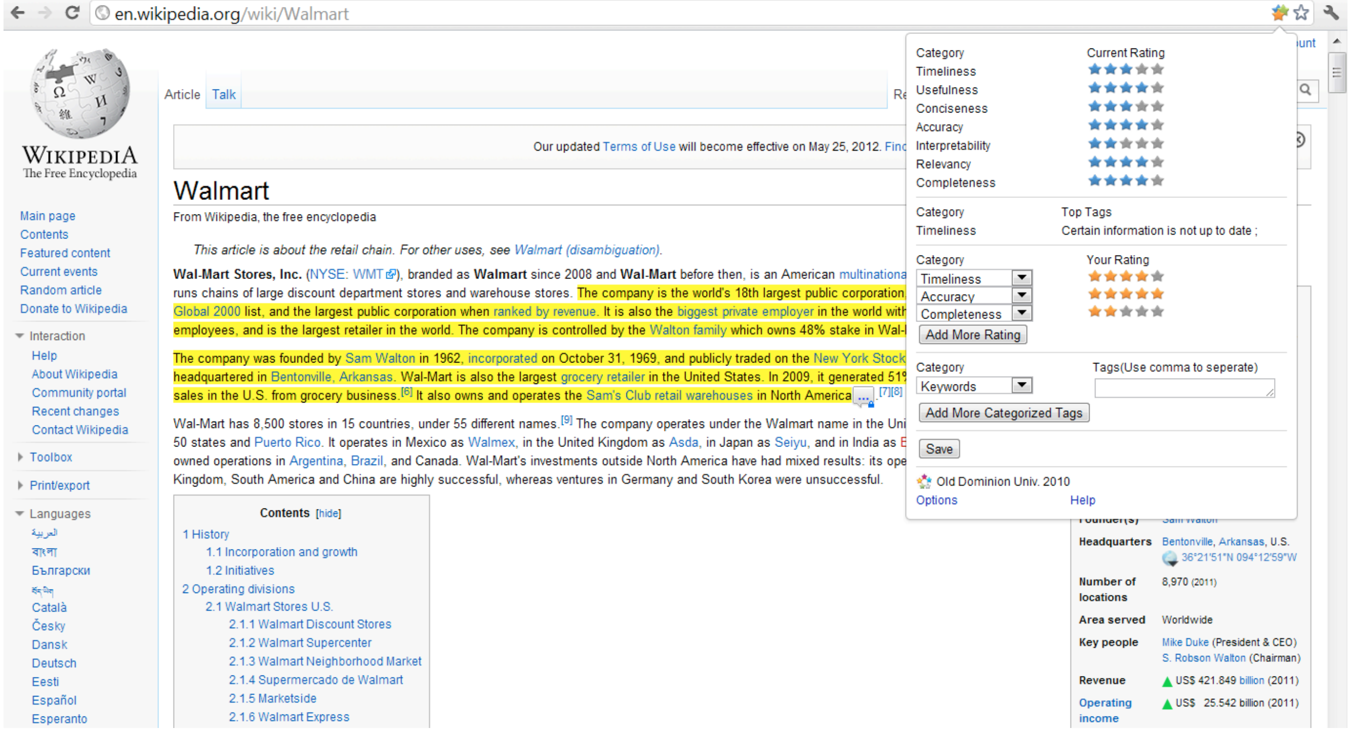

Figure 5. Prototype 2: icon with pop-up window on mouse-over.

For both prototypes, we have implemented in-situ annotation to collect and display detailed feedback about any fragment within a Web page (Figure 6). This level granularity is desirable when user feedback is used to diagnose problems and continuously improve IQ. Highlight can be turned off to avoid interference with the reader.

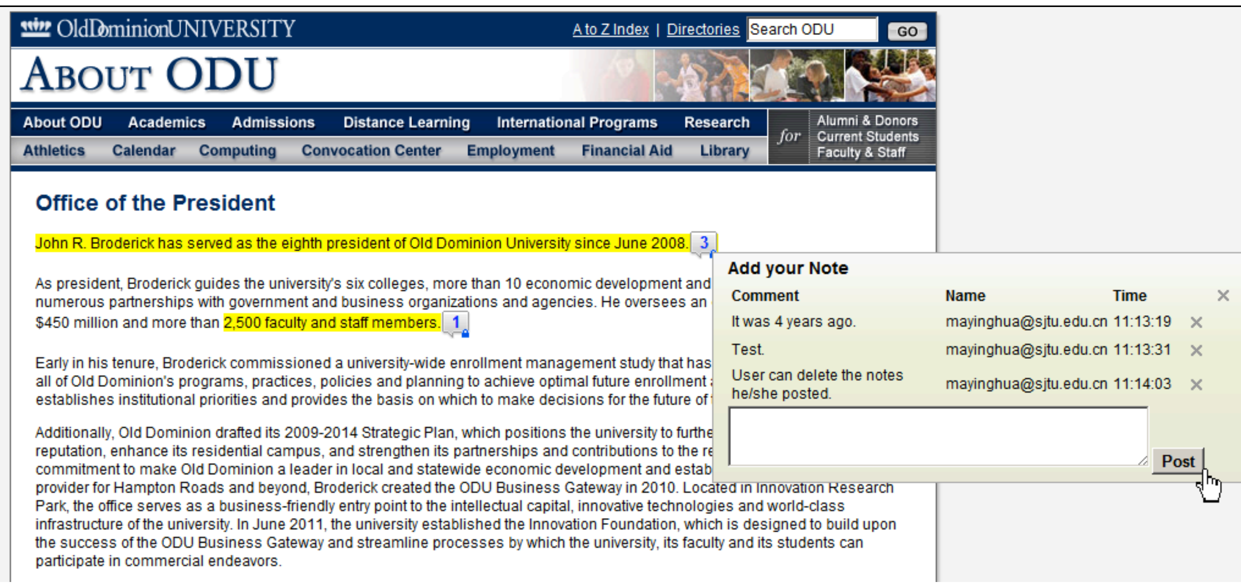

Figure 6. In-situ annotation for specific IQ feedback.

EVALUATION

After a pilot testing of both prototypes with a small group of users, we chose to use prototype 2 in an evaluation study to evaluate its effectiveness and gather user feedback. MBA students in an IT Strategy course are asked to install the tool and use it with at least two reading assignments for a firm strategy analysis project: (1) a Wikipedia article about Wal-mart, and (2) any web pages the student reads for completing the project. A questionnaire containing 18 questions is also supplied to students to collect additional feedback.

The evaluation is still in process and we can make a few observations abut user inputs received so far. There seem to be consensus among users in intrinsic quality dimensions, however, the variations in certain contextual dimensions are large. Some users added their own dimensions. Text highlighting is useful, but some users desire to have a feature to switch the highlights on and off. More numeric ratings than non-numeric comments have been added by users, which indicates that users prefer to use simple methods to supply feedback. Users are also more inclined to using annotation feature than the non-numeric comments (tagging) feature.

We plan to run another evaluation test in Fall 2012. In addition to the two tasks given in Spring 2012, we plan to add a third task: students will be asked to read an instructor-created web page that contains several purposely introduced “mistakes”. This will allow us to see whether the tool is usefully for users to collaboratively identify the errors and provide useful feedback for quality improvement.

CONCLUSION AND FUTURE RESEARCH

We have identified a critical deficiency of existing methods for assessing IQ on the Web and proposed a collaborative approach to addressing the deficiency. As a critical first step towards implementing the solution, we have developed two prototypes of the user interaction component.

Our goal is to implement the entire solution approach in the near future. Currently, we are evaluating the user interaction component with a pilot usability tests that involve focused user groups. The improved component will be deployed as a tool to collect field data. The IQ metadata collected will be used as a testbed to support the development and tuning of the machine-based algorithms.

REFERENCES

1. Adler, B.T., and Alfaro, L.d. "A content-driven reputation system for the wikipedia," in: *Proceedings of the 16th international conference on World Wide Web*, ACM, Banff, Alberta, Canada, 2007, pp. 261-270.
2. Bovee, M., Srivastava, R.P., and Mak, B. "A conceptual framework and belief-function approach to assessing overall information quality," *International Journal of Intelligent Systems* (18:1) 2003, pp 51-74.
3. Dasu, T., and Johnson, T. *Exploratory Data Mining and Data Cleaning* John Wiley & Sons, Inc., 2003, p. 203.
4. Doan, A., Ramakrishnan, R., and Halevy, A.Y. "Mass Collaboration Systems on the World-Wide Web," *Communications of the ACM*) 2010 (forthcoming).
5. Eysenbach, G., Powell, J., Kuss, O., and Sa, E.-R. "Emiprical studeis asessing the quality of health information for consumers one wh World Wide Web," *JAMA* (287:20) 2002, pp 2691-2700.
6. Fan, W. "Dependencies revisited for improving data quality," in: *Proceedings of the twenty-seventh ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, ACM, Vancouver, Canada, 2008, pp. 159-170.
7. Herzog, T.N., Scheuren, F.J., and Winkler, W.E. *Data Quality and Record Linkage Techniques* Springer Publishing Company, Incorporated, 2007, p. 234.
8. Klein, B.D. "The Detection of Data Errors in Computer Information Systems: Field Interviews with Municipal Bond Analysts," *Information Resources Management Journal* (13:3) 2000, pp 23-32.
9. Lee, Y.W., Strong, D., Kahn, B., and Wang, R.Y. "AIMQ: A Methodology for Information Quality Assessment," *Information & Management* (40:2) 2002, pp 133-146.
10. Madnick, S., and Wang, R.Y. "Introduction to total data quality management (TDQM) research program," TDQM-92-01, MIT Sloan School of Management, Cambridge, 1992.
11. Madnick, S.E., Wang, R.Y., Lee, Y.W., and Zhu, H. "Overview and Framework for Data and Information Quality Research," *ACM Journal of Data and Information Quality* (1:1) 2009, p Article #2.
12. Mandl, T. "Implementation and evaluation of a quality-based search engine," in: *Proceedings of the seventeenth conference on Hypertext and hypermedia*, ACM, Odense, Denmark, 2006, pp. 73-84.
13. Miller, G.A. "The magical number seven, plus or minus two: Some limits on our capacity for processing information," *Psychological Review* (63:2) 1956, pp 81-97.

14. Orr, K. "Data Quality and Systems," *Communications of the ACM* (41:2) 1998, pp 66-71.
15. Price, R., and Shanks, G. "A Semiotic Information Quality Framework: Development and Comparative Analysis," *Journal of Information Technology* (20:2) 2005, pp 88-102.
16. Stvilia, B., Gasser, L., Twidale, M.B., and Smith, L.C. "A Framework for Information Quality Assessment," *Journal of the American Society for Information Science and Technology* (58:12) 2007, pp 1720 -1733.
17. Wang, G.A., Jiao, J., and Fan, W. "Searching for Authoritative Documents in Knowledge-Base Communities," in: *30th International Conference on Information Systems (ICIS'09)*, Phoenix, AZ, 2009.
18. Wang, R.Y., Lee, Y.W., Pipino, L.L., and Strong, D.M. "Manage Your Information as a Product," *Sloan Management Review* (39:4) 1998, pp 95-105.
19. Wang, R.Y., and Strong, D.M. "Beyond accuracy: What data quality means to data consumers," *Journal of Management Information Systems* (12:4) 1996, pp 5-33.
20. Zeng, H., Alhossaini, M.A., Fikes, R., and McGuinness, D.L. "Mining Revision History to Assess Trustworthiness of Article Fragments," in: *International Conference on Collaborative Computing: Networking, Applications and Worksharing*, IEEE, Atlanta, GA, 2006.